Evaluating the Usage of Museum Digitized Collections

Emily Horbatch

ehorbat1@jhu.edu

Johns Hopkins University, Museum Studies Digital Curation Certificate Program

Fall 2017

December 10, 2017

Revised: July 29, 2018

Abstract

Driven by perceived value, resources often are allocated to the digitization of museum collections. But what evidence supports this value? Calls for proof of digitized collections value may be met by data derived from a usage study. A usage study captures data relating to public consumption of a given dataset via implemented methods which collect related quantitative and qualitative statistics. These statistics represent reach and impact, respectively defined as number of viewers and instances of content use. In its capacity to reveal public interests, this usage data can be exploited to enact digitized collections change which best serves users. This change cultivates relevance, which supports the sustainability of digitized collections. Challenges may arise within the process of a usage study, but problems are navigable. These identified themes collectively indicate the practicality of usage studies of digitized collections and the advantageous nature of usage data ultimately to guide museum professionals in their own usage study experiences.

Table of Contents

Introduction

In attempts to maintain relevance in an increasingly digitally-reliant world, museums turn to the production of digitized collections to relate to their audiences. But, do digitized collections have value? This is an important question when the processes behind digitization and making digitized collections accessible constitute investments of museum resources. It can be beneficial for museums to determine whether their digitized collections represent an effective use of time and funds, and this may be achieved by understanding the usage of these digitized collections. This paper explores how the usage of museum digitized collections can be evaluated to find value.

An evaluation of the usage of museum digitized collections may be achieved by assessing reach and impact, as reported in an article titled "Measuring the Impact of Museum Websites" by Eric Holter (2017). "Reach" is a quantitative measurement of breadth of access (Holter, 2017). This a revealing indicator of efficiency, but numbers alone offer an incomplete picture of digitized collections usage. "Impact" is an additional metric by which digitized collections usage may be judged. "Impact" represents a qualitative indication of "[…] the effect or influence that one agent, event or resource has on another" (Ball & Duke, 2015, p. 3). Thus, recognition of both reach and impact is important in responding to the above problem statement, and as such related sub-questions guide the direction of this paper:

- What methods exist to evaluate the reach and impact of museum digitized collections?
- What challenges impede the implementation of methods which capture reach and impact of museum digitized collections?
- What are the implications of the information generated via these tools? Are there case studies of institutions which have monitored the usage of their digitized collections?

Research Design

To investigate how museums can evaluate the usage of their digitized collections, research was conducted via the methods below to provide the following data:

- Resources which explore methods by which reach and impact can be measured were compared in a meta-analysis. This meta-analysis permitted an understanding of common means to evaluate usage of digital content, and associated challenges.
- Published case studies were reviewed to identify measures taken by specific institutions to monitor the usage of their digitized collections. Attention was paid to the data yielded by these usage measures, and the implications of this data for these particular institutions.

The above research method decisions were made through a preliminary literature review. With library usage study-related articles highly represented within search engines and library databases, this preliminary literature review supported the need for this type of museum-related research. A dependence on related literature was necessary both to identify the breadth of available usage methods, and to capture the experiences of various institutions with usage tracking. Analysis of the usage study approaches and experiences described within this literature offer insight to guide professionals as they make digitized collections decisions.

Literature Review

A literature review revealed relevant museum, library, and archive-related resources by which digitized collections usage studies could be explored. These resources were gathered through a search of Google, Google Scholar, and two library databases, JSTOR and the Johns Hopkins University Sheridan Libraries.

Various resources express the relationship between usage studies and the sustainability of cultural heritage institutions and their digital initiatives. The recognition of this relationship by these institutions is evident within several case studies. These case studies describe efforts taken by specific institutions to track the usage of their digitized collections and to respond to user needs and interests indicated by resulting usage data.

The article "Measuring the Impact of Museum Websites" permits recognition of the need to capture data pertaining to both reach and impact to grasp a complete glimpse of usage. Other resources collectively offer a broad look at the scope of a usage study. Resource information spans from identification of reach evaluation methods which provide quantitative statistics (e.g. Google Analytics; Log File Analysis), to description of qualitative methods which indicate impact (e.g. Content Analysis; Focus Groups), to types of data collected by these methods, such as page views, which are a quantitative metric. These resources consider probable challenges of measuring reach and impact and potential barriers to responding to usage data, indicating that usage evaluation methods and resulting usage data grapple with the unique circumstances of every institution.

These resources offered the insight needed to explore the questions posed. Insights from these resources collectively offer guidance for institutions which seek to evaluate the usage of their digital content.

Evaluating the Usage of Museum Digitized Collections

Museum digitized collections may be a standard response to public interest in digital content, but is this response an adequate one? When "[…] Web presence does not necessarily guarantee the use of digital resources" (Mills, 2015, p. 161), a usage study can reveal the success of this response via methods which yield data pertaining to the digitized collections actions of users. Such studies have relevance when museum budgets call for demonstrations of program impact (Bertacchini & Morando, 2013, p. 70). In the face of this pressure, tangible evidence can offer the confidence and understanding that speculation cannot (Ball & Duke, 2015, p. 2). Usage data can support informed change when data reveals instances where greater use could be attained (Biswas & Marchesoni, 2016, p. 19). Thus, as digitized collections vie for the attention of users, usage data can have profound implications for digitized collections sustainability. This report considers all of this, and references various museum, library, and archive-related resources to offer an in-depth look at the process of implementing a digitized collections usage study, including methods which evaluate usage, implications of usage data, and challenges which may arise in the midst of a usage study.

Reach and Impact

Prior to conducting a usage study, it can be helpful to understand that a complete picture of usage requires an assessment of both reach and impact. Any museum program expected to be impactful first must be reachable (Holter, 2017). However, reach and impact may be differently judged, as "[…] the number of people impacted will always be a small subset of those we reach" (Holter, 2017). For example, while content with a high number of page views may be determined to have heavy reach, a singular instance of clear impact may be considered a success (Holter, 2017). To branch off of this, an assessment of the reach and impact of a given program first

requires a museum to consider what the evidence of the reach and impact of this program will look like, and then select relevant tools to gather this data (Hull, 2011, p. 26). This understanding will avoid an instance when "'[…] people gather lots of data in an analytics platform and then look for a problem to solve, but that will only threaten your return on investment'" (Everett, 2017, p. 20).

Tracking Usage

As prior indicated, visitor numbers (reach) and instances of inspiration (impact) both offer valuable insight. Implementation of quantitative and qualitative methods to capture both reach and impact can offer a complete picture of museum digitized collections usage.

While the Toolkit for Digital Scholarly Resources (TIDSR) is not a method in itself, it is an invaluable resource for any institution with plans to implement a usage study. TIDSR directs digital information professionals toward quantitative and qualitative methods which can be individually or collectively employed to measure the usage of digital programs (Oxford Internet Institute, University of Oxford, n.d.). This toolkit identifies many of the methods which are briefly described below. These methods are categorized into their respective quantitative and qualitative groupings. Considering the "[…] wide variance in what they measure and some differences in how they measure it" (Ball & Duke, 2015, p. 12), an understanding of the function of each of these methods can be helpful to the selection process. This method selection ultimately is based on what a museum hopes to gain from a usage study.

Quantitative Methods

The methods described below may not be all-inclusive, but are examples of those which express reach via the numeric statistics they track, and thus can be categorized as quantitative.

Analytics can offer "[…] statistics about traffic and visitors to your website" (Meyer, n.d. g). Of the various analytics tools by which these statistics can be gauged, Google Analytics may best be known, as statistics point to the worldwide popularity of this tool (Farney, 2016, p. 5). Google Analytics offers three distinct categories of data: audience reports collect information about the user, including user geographic location and usage duration; acquisition reports gather information related to how users are finding the resource, whether by search engine or other web location; behavior reports indicate user actions within the resource with such data as page view counts (Szajewski, n.d.). Event tracking and filters account for the accuracy of analytics data, for example by instructing Google Analytics to refrain from recording the activities of known spam websites (whose activities do not represent actual users) (Farney, 2016, p. 6). Google Analytics may be further customized via Google Tag Manager, which "[...] is a script management system that […] can easily track events and custom dimensions and metrics for Google Analytics" (Vogl, Zhou, Draper, & Shelstad, 2016, p. 32).

Bibliometrics relates to a category of tools which function to find citations which reference certain digital information. Google Scholar Citations is one instance of bibliometrics in action. Google Scholar Citations reviews a wide breadth of materials, including "[…] informal scholarly communication (such as presentations and conference papers) […]" (Meyer, n.d. c), to capture citations that may remain unseen by other bibliometrics software. Google Scholar Citations present this information in a graphical manner, permitting citation counts to be grasped without having to count each citation instance (Google, n.d.).

Log file analysis tracks instances of web page access and file downloads. This method may most accurately indicate usage via file download counts because "[…] it implies a desire to look at the actual data […]" (Ball & Duke, 2015, p. 5). When a web page is requested, a web

server sends the requested page to the user. With the implementation of web log file analysis software, the web server creates a log file for each request, within which it may note such information as web page URL, time of request, and the username of the individual requesting the URL (Eccles, n.d.). The implemented software may then translate these log files into "[…] graphs to illustrate the most popular pages in a site, where the requests for the pages in the site came from, and graphs to show changes over time […]" (Thelwall, n.d.).

Surveys similarly can yield quantitative usage information. As an example, a survey could ask participants whether they have ever accessed a museum's digitized collections. While a survey can be offered to visitors on site, various websites offer platforms for museum professionals to generate surveys with a wider reach. Considering the unique context of every museum, prior to selecting a survey website, the capacities of these survey websites may need to be evaluated to understand whether they serve museum needs (Meyer, n.d. f).

Qualitative Methods

With usage quantitatively determined, qualitative methods can identify how information from museum digitized collections is being used. This impact may be gauged through any of the qualitative methods recognized below, which again may not be all-inclusive.

Content analysis is one qualitative method with the capacity to capture impact. This method first requires searches within Google, Wikipedia, and/or other search engines for materials which reference the digitized collections. Bibliometric tools automate this search process. Various qualitative data analysis software kits exist to parse through these found resources on behalf of museum staff to reveal major insights and themes (Meyer, n.d. d).

Feedback opportunities can pinpoint digitized collections impact. Digital venues enable direct communication between museums and the public. Social media platforms have been

embraced by the museum world for their capacity to "[…] spread museum knowledge (Cameron 2008) and receive knowledge back from visitors (Russo 2011)" (Gerrard, Sykora, & Jackson, 2016, p. 233). Social media permits freedom to express thoughts and concerns, and contact pages on museum websites can perform the same function. On these social media platforms, impact may be expressed both explicitly and implicitly. Social media users explicitly may indicate whether museum digitized collections have had an impact via comments made within posts. Social media websites may facilitate indications of implicit impact, as posts may be "up-voted" on Reddit, or "retweeted" on Twitter (Ball & Duke, 2015, p. 6).

One-on-one interviews and multi-person focus groups similarly require direct contact to encourage the public to share about their digitized collections experiences. Both instances involve asking digitized collections stakeholders various reach and impact-related questions, and may be most revealing if stakeholders represent diverse backgrounds (Mills, 2015, p. 164). Focus group discussions may offer a distinct advantage over one-on-one interviews as participants elicit information from one another as they question and respond to the statements of others (Meyer & Eccles, n.d.). In either respect, these communication methods can reveal information which otherwise may remain hidden.

Impact measurement services capture impact metrics on behalf of an institution. As an example, PlumX is a subscription-based tool which measures impact by surveying various web locations to gather "[…] captures: the number of times the resource has been marked as being of interest […] mentions: the number of blog posts written about it, the number of comments made about it (on Facebook, Slideshare, YouTube, etc.), the number of reviews received (on Amazon or Goodreads)" (Ball & Duke, 2015, p. 9). Social media comments which reference a given resource suggest interest and understanding, so it is noteworthy that PlumX pays particular

attention to social media as a source of usage data. This social media attention can be seen in other data that this tool collects, like "[…] the number of times the resource has been recommended (e.g. by means of 'likes' on Facebook, '+1s' on Google+, net upvotes on Reddit, tweets)" (Ball & Duke, 2015, p. 9). PlumX presents collected data within tables and diagrams (Ball & Duke, 2015, p. 9).

Method Challenges

Although the above methods have the capacity to capture usage, implementation of these methods may prove challenging. This may not necessarily be a reflection on the methods themselves, but of the context of a museum, its digitized collections, and its users. The following challenges should be kept in mind as usage evaluation methods are selected and implemented in attempts to execute a museum digitized collections usage study.

Google Analytics requirements can pose a challenge to implementation of this usage evaluation method. To enable usage tracking on a given digitized collections page, Google Analytics requires its HTML coding to be embedded into the page. A museum which aspires to track the usage of the entirety of its digitized collections would need to devote significant time to integrate this coding into each digitized collections page (Meyer, n.d. b). Also, Google stores the information that it tracks on its own servers. This may be problematic when security is a concern, such that museum policies "[…] prevent data being stored on an outside server […]" (Meyer, n.d. b). In light of this, Google Analytics code embedding may not be an option for some museums.

Digitized collections design can influence the accuracy of usage data. Google Analytics will track a single page view as multiple when several frames pop up within a web page (Oxford Internet Institute, University of Oxford, 2013, p. 3). Digitized collections navigation systems

similarly can create a flawed picture of usage when users have minimal control over the web pages that they visit. This is reasserted by the Oxford Internet Institute, University of Oxford (2013): "Having a navigation system that causes users to repeatedly visit one particular page (such as search) can lead to a very lopsided view of how the site is used […]" (p. 3).

Bibliometrics may not be able to fully capture the extent of digitized collections usage when confusion surrounds citations. This confusion may be due to a lack of citation guidance both within the digitized collections itself and citation style guides: "Many of the style guides do not have clear guidance for how to cite a database, for instance, or whether to cite a digitised resource in a way that identifies its digital location, or that cites the original item, whether or not the researcher actually consulted it" (Meyer, n.d. h). Also, bibliometrics can only find citations so long as there are "[…] persistent, unique identifiers to assist with unambiguous referencing […]" (Ball & Duke, 2015, p. 5). Inattention paid here may either result in user failure to cite a digitized collections material or an inaccurate citation if a user is citing the physical item rather than its digitized copy (Meyer, n.d. h). In either of these circumstances, bibliometrics will be unable to find related citations. Digitized collections which offer pre-made citations for users to employ may be able to skirt around this issue and at the same time avoid copyright concerns (Meyer, n.d a). As an aside, it takes time for resources to be published, so reach and impact may not be seen immediately within published materials (Ball & Duke, 2015).

While surveys can be a means to understand usage, museum professionals must consider that users may not always be eager to answer questions. This situation can be compounded by survey length, and the mistrust which surrounds sharing information online in light of the potential for misuse. This mistrust may be particularly problematic when a museum creates a survey on a website that does not bear the institution's name (Meyer, n.d. f). Considering these

concerns, this method may not always offer a complete picture of usage, but the Toolkit for Digital Scholarly Resources does offer several tips to coax survey responses: "Send out up to 3 reminders. […] Make sure you stress that this is academic not market research. […] Include an (accurate) indication of how long the survey will take – an inaccurate indication can really alienate participants" (Meyer, n.d. e).

Public feedback may not be an absolute indication of digitized collections reach and impact. Participant responses within surveys and focus groups may not represent the truth. When digitized collections-related social media posts are shared by the public, this may not necessarily indicate genuine reach or impact, but instead may represent vague interest, or even a means to achieve 'likes' or 'favorites' (as dependent on the social media platform). Ball & Duke (2015) comment on this issue: "It is arguable, for example, what proportion of the tweets about a resource represent real engagement with it, and what proportion are simply passing on an interesting title" (p. 12). Also, considering that not all digitized collections users have social media accounts, not all will post about their digitized collections experiences (Ball & Duke, 2015, p. 6). Similarly, digitized collections may have existed prior to the dawn of social media, so instances of digitized collections-related social media posts may be an incomplete representation of the extent of reach and impact (Ball & Duke, 2015, p. 13).

Finally, the funding required to evaluate reach and impact can be a barrier for museums. Digitization initiatives often are funded in the short term. This can be problematic when usage studies require long-term assessments of reach and impact to capture a realistic glimpse of digitized collections usage (Hughes, 2012). Time can offer perceptions of the greater ebbs and flows of digitized collections usage via comparisons of data from past months with present data. As Meyer (n.d. g) similarly states, "In this way, you can see whether you are experiencing big

changes, are getting traffic from new sources, are seeing much more interest in particular pages, or other anomalies that will often need further research to explain."

Responding to Usage Data

Following the selection and implementation of quantitative and qualitative usage evaluation methods, a review of accumulated data serves to generate an understanding of the state of digitized collections' reach and impact. Usage data may indicate the need to readdress certain variables to promote digitized collections engagement via any of the actions listed below.

Page view data identified by quantitative methods can indicate user interests. High view counts for given pages may infer interest in these pages, and as such reinforce the continued digitization of related materials. The Ball State University Digital Media Repository acted in response to high page view counts by adding materials related to the subject areas of these heavily viewed pages (Szajewski, n.d.). On the other hand, low counts for a page may "[…] inform decisions about data retention" (Ball & Duke, 2015, p. 3). Or a response to low page view counts could follow the example of the Health Sciences Libraries of the University of Minnesota, which reviewed its Google Analytics data to "[…] inform decisions to feature underused pages more prominently […] subsequent follow-up data analysis showed that the redesign achieved its aims" (Szajewski, n.d.).

Low reach and impact statistics may prompt a reassessment of restrictions on digitized collections access. Ball & Duke (2015) consider the correlation between impact and access: "[…] for maximum impact such restrictions should be the minimum necessary to comply with legal, ethical and other obligations" (p. 15). Nonrestrictive access heightens the relevance of digitized collections (Leon, 2012). With this in mind, removing subscriptions and other unnecessary restrictions on digitized collections access may be one means to bolster reach and

impact. Such a move also supports the responsibility of museums to share cultural heritage (Leon, 2012). Considering that digitized collections serve to introduce users to objects within the museum's collections, another effect of this expanded access may be increased physical visits as users venture to museums to view the tangible manifestations of digitized objects (Bertacchini & Morando, 2013, p. 65). Access to digitized collections similarly can heighten the impact of physical visits as users retrieve information from digitized collections via mobile devices: "This system can even enhance the quality of the user's experience by allowing for a more thorough inspection of an artwork's details through ultra-high-resolution images […]" (Bertacchini & Morando, 2013, p. 64). Thus, fewer access restrictions can set the stage for expanded reach and impact as greater numbers are permitted to interact with digitized collections content.

A museum may reassess discoverability in response to digitized collections usage data. Discoverability is key when institutions are vying for the attention of web users. This is echoed by Szajewski (n.d.), who notes that "[…] digital archives are competing with a plethora of online materials and information for the interest and visitation of users […] Thus, emphasis on discoverability is not only beneficial but essential in assuring the cultural relevance of digital archives." A museum which expects its digitized collections to be accessed should understand that users first must be aware of its existence. Considering that discoverability is key to usage, museums may respond to minimal usage counts via promotion of the digitized collections. Museum social network platforms can perform this promotion function with posts which tempt users into clicking a link to the digitized collections (Breeding, 2009, p. 33).

Search engine query data captured by Google Analytics can be used to amend digitized collections metadata. This search engine query data identifies the keywords typed into search engines by which users are finding the digitized collections (Szajewski, n.d.). Review of these

keywords can be key to understanding the language of users, enabling those managers of digitized collections "[…] to learn how to describe assets in ways that makes them discoverable to a greater number of potential users" (Szajewski, n.d.). Search keywords added as tags to relevant object records within the digitized collections may serve as a type of search engine optimization.

The social media share-ability of digitized collections may be considered in response to reach and impact data. Users can serve an important, often unsolicited role in digitized collections promotion. Considering current public interest in social media, and thus the visibility of social media posts, share "buttons" added to object records may increase the exposure of digitized collections as these buttons facilitate social media sharing. Social media visibility may be corroborated by Google Analytics data which identifies how users are finding the digitized collections (Szajewski, n.d.). Biswas and Marchesoni (2016) comment on this visibility as they reference the findings of a certain usage study: "Ladd's findings reveal that few users came to the main digital collections website to search and browse; instead, most arrived via external sources such as […] social media sites" (p. 20). Added benefits of the social media share-ability of digitized collections may be indications of reach and impact: "If people are moved to share or discuss a dataset with friends, colleagues and the wider world, there is a likelihood it has affected them in some way, meaning it is worth looking closer for evidence of impact" (Ball & Duke, 2015, p. 6).

Digitized collections impact may be found in the ability of users to add information to object records. Users may have knowledge about the provenance of a given object or be able to find connections between digitized collections objects. This knowledge could be shared via a Web 2.0 function for annotations on each page of the digitized collections. Museums which

encourage users to share what they know may not only find expanded object meaning but greater user investment in the digitized collections. This is repeated by Bertacchini and Morando (2013) in the specific context of digitized collections of artworks: "As noted by Eschenfelder and Caswell (2010), adopting strategies that increase […] reuse of collections could engage amateur experts in assisting with descriptions and the addition of context to artworks, thereby increasing the value of collections and public commitment to artworks" (p. 69).

Insights gained during direct communication with digitized collections stakeholders can foster awareness of user interests. As an example, similar to a question posed within a 2008/2009 Continuous Household Survey conducted for museums in Northern Ireland, a survey may ask respondents to identify that which would encourage them to make more frequent use of the digitized collections (Hull, 2011, p. 7). Feedback may aid digitization efforts if respondents indicate that they would like to see a certain artist or medium better represented within the digitized collections. This could prompt the enactment of a program which considers user requests for digitization. With user input, "[…] institutions are sure to digitize materials that will directly impact users, and consequently, the wider research community" (Mills, 2015, p. 164).

Partnership with aggregator websites may be a response to digitized collections reach and impact statistics. Any access may be considered good access, regardless of where it happens (Biswas & Marchesoni, 2016, p. 20). Aggregators offer both increased visibility and connectedness beyond a museum's capacities on its website and social media pages. While this visibility may bolster reach, connectedness may generate meaning and thus impact: "As most collections represent only part of the corpus of any single artist, subject area or era, the need to pull together cultural resources from across many institutions may be seen as an intellectual imperative for enhancing users' experience of museum collections (MTM London, 2010;

Tanner, 2004)" (Bertacchini & Morando, 2013, pp. 63-64). Partnerships with aggregators may have the added benefit of use of their analytics systems so that museums can ascertain usage on these aggregator websites (Biswas & Marchesoni, 2016, p. 21).

Response Concerns

While it can be important to respond to usage data, as exemplified above, staff may disagree about how to move forward. The following scenarios are drawn to prepare museum professionals for discord as they attempt to respond to digitized collections usage data.

A program developed in response to usage data may prove challenging to implement when staff express discordant opinions about this course of action. As an example, certain individuals may be concerned about the potential for inaccuracy if users are permitted to annotate digitized collections object records (Bertacchini & Morando, 2013, p. 69). As another example, some may find it imperative to increase access to and visibility of digitized collections while others remain concerned about retaining control over digitized collections (Bertacchini & Morando, 2013, p. 70). When scenarios like these are encountered, it can be important to work through differences in opinion to reach some sort of greater understanding or consensus. For example, to appease concerns about inaccurate user annotations, a compromise may be a comment section offered on each object record page for users to submit annotation requests, which then could be approved prior to "official" annotation.

The perceived correlation between greater access and decreased control inferred in the prior paragraph raises legitimate concerns amongst museum professionals. Increased access to digitized collections may be viewed as a disruption to the museum "[…] position as gatekeepers of authoritative and trusted digital content" (Bertacchini & Morando, 2013, p. 60). Aggregators and other parties infringe upon this gatekeeper position when museums hand over digitized

collections control. In spite of this concern, "If cultural heritage organisations do not expose data in ways that digital natives want to use it, they risk becoming irrelevant to the next generation" (Verwayen, Arnoldus, & Kaufman, 2011, p. 4). A solution to placate concerns about loss of control over digitized collections may be to solely offer low-resolution images on the web so that any user looking for a "master" image for publication would need to contact the museum directly (Bertacchini & Morando, 2013, p. 65). Another related concern may be elimination of sources of revenue generated via a fee to access the digitized collections. However, some museums may not find it pertinent to require such a fee when "[…] the costs of disseminating […] information have decreased, the transaction costs associated with charging for access to information and controlling redistribution have formed a barrier to access in themselves" (Bertacchini & Morando, 2013, p. 64). Also, success exists in forms other than monetary, as reach and impact statistics indicate (Verwayen et al., 2011, p. 4).

A dependency on user input may draw similar staff concerns. Communication with the public can reveal their perceptions of the digitized collections, and thus serve to support comprehension of low usage numbers. However, user input may be non-representative of the needs and interests of all digitized collections stakeholders. Thus, consideration of user input may risk non-advantageous digitized collections changes. These concerns may be appeased by reaching out to a variety of individuals for input, and then reviewing this input within the context of the institution's mission and aims (Mills, 2015, p. 167). This discretion holds greater significance in the face of "[…] copyright, privacy, or other restrictions impeding their digital conversion" (Mills, 2015, p. 165-166), which may not be considered by users as they request items for digitization.

These concerns are valid, but responses to usage data equally are justifiable. Verwayen et al. (2011) suggest "[…] courage to take some necessary risks and a strong commitment to the mandate of the cultural heritage sector, which is to enable society to realise the full value of the cultural legacy that is held in the public realm" (p. 4).

Usage Study Examples

The following two case studies detail the digital collections usage study experiences of specific cultural heritage institutions, including methods implemented, data gathered, and data implications.

Derby Museums, Derby, UK

While this case study describes attempts by Derby, United Kingdom museum organization Derby Museums to identify instances of inspiration tied to its public events, insight nonetheless can be gained from this case study as museums seek evidence of digitized collections impact. Impact has comparisons to inspiration, defined within this case study as "[…] an experience, or set of experiences, combining rational thoughts and emotions, resulting in the expression or enactment of fresh ideas" (Gerrard et al., 2016, p. 233).

While the quantitative and qualitative methods mentioned at the outset of this report may be implemented to capture digitized collections usage data, museums with available resources may design usage evaluation methods to fulfill their specific needs. This case study details actions taken by Derby Museums to develop an automated program to review public social media posts for notions of inspiration tied to two distinct museum events (Gerrard et al., 2016, p. 234). The value of social media for an inspiration study is apparent in the following conclusion made by Gerrard et al. (2016): "[…] a positive factor that emerged from using Twitter was the ease with which smartphones could be used to Tweet: this meant that a substantial proportion of

the data was created during the two events, therefore potentially at the point where visitors may have actually felt inspired" (p. 236). The decision to develop an automated program to assess inspiration both was considered necessary in light of the quantity of related social media posts, and attainable via algorithms which exist to recognize emotion within text (Gerrard et al., 2016, p. 234). Thus, the program built upon existing lexical database FrameNet, which can "[…] interpret potentially complex experiences and situations (Fillmore 1976)" (Gerrard et al., 2016, p. 235).

Social media platform Twitter was chosen as the focus of this inspiration study for three distinct reasons: "(i) Twitter's API provided easy access to sample data by searching for hashtags; (ii) the data was in the public domain […]; and (iii) the organisers encouraged visitors to Tweet about their experiences before, during and after attending the events" (Gerrard et al., 2016, p. 236). These events each spanned for a few days in 2014, one the Derby Mini Maker Faire during which creators showed their works, the other MuseoMix UK 2014, an event which enabled participants to design exhibitions with objects from the Derby Museums collections (Gerrard et al., 2016, p. 237). Those who conducted this study compiled tweets relating to these distinct museum events via museum promoted hashtags (e.g. #DMMF14; #MMUK14) and event name mentions (Gerrard et al., 2016, p. 237). FrameNet then parsed through the tweets gathered to find instances of inspiration-related lexical units, or terms, within these tweets (Gerrard et al., 2016).

This Derby Museums inspiration evaluation method yielded the following data: 24% of Maker Faire tweets and 22% of MuseoMix tweets contained inspiration-related lexical units, as determined by FrameNet (Gerrard et al., 2016, p. 239). Of these respective percentages, tweets most often contained lexical units related to creating: "The potential evidence of inspiration

found in Twitter data collected for this research predominantly belonged to types 4 and 5: people trying out new creative techniques or brand-new activities" (Gerrard et al., 2016, p. 241). Considering the focus of Derby Museums on promoting creativity, these results were promising.

In light of data collected, certain conclusions can be made about the capacity of social media to indicate inspiration, and FrameNet to identify it. Upon a review of tweets found by FrameNet to contain inspiration-specific lexical units, Derby Museums staff for the most part agreed with these FrameNet findings (Gerrard et al., 2016, p. 243). However, FrameNet failed to recognize certain tweets that staff considered to be inspiration-related, which may have been a result of staff being "[…] generous in labelling Tweets as 'containing expressions of inspiration'" (Gerrard et al., 2016, p. 244). The text-based nature of FrameNet also may have played a role in this, as this lexical database is unable to identify instances of image-based inspiration (Gerrard et al., 2016, p. 245). The 140-character limit imposed on Twitter users may have inhibited their ability to adequately express inspiration. The reliance upon event hashtags and mentions of event names to retrieve tweets may have omitted tweets with different or without hashtags, but which nonetheless related to inspiration (Gerrard et al., 2016, p. 236). While the pressure to tweet with these hashtags may have produced tweets that reference inspiration, these tweets may not be true depictions of inspiration: "There is a gap between expressions in written language and other online behaviours compared to the true experienced personal states of the individuals producing the social media content […]" (Gerrard et al., 2016, p. 247). Finally, considering the short-term nature of museum events, and that social media serves as a means for the public to immediately react to these events, it may be challenging to find inspiration-related tweets after an event has concluded to understand the long-term impact of these events.

Derby Museums suggests the following actions to combat some of the prior addressed concerns. Considering that retweet redundancy can paint an inaccurate picture of inspiration, the Derby Museums omitted retweets from its inspiration analysis (Gerrard et al., 2016, p. 238). However, in light of the other factors which may come into play to inaccurately portray inspiration, Derby Museums recommends that social media analysis be one of several methods implemented to capture inspiration (Gerrard et al., 2016, p. 248). Finally, considering the limited ability of users to express inspiration via Twitter, Derby Museums suggests that a social media platform which permits a larger number of characters per post may be better suited for such a study (although, Twitter since has doubled the number of permitted characters per post) (Gerrard et al., 2016, p. 247).

Hunter Library, Western Carolina University, Cullowhee, NC

The following case study reviews actions taken within a usage study performed on the digital collections of the Hunter Library of Western Carolina University. While this usage study happened outside of a museum context, actions taken within the process of the study exemplify much of what has been discussed within this report.

The Hunter Library conducted a 27-month usage study of its 14 digital collections with quantitative tool Google Analytics (Biswas & Marchesoni, 2016, p. 19). While conducting this study, Hunter Library professionals paid close attention to Google Analytics data relating to who digital collections users were (user locations and repeat visitors) and how users found the digital collections (search keywords and referrals) (Biswas & Marchesoni, 2016, p. 21). This data was considered most telling, given the aim of this usage study to identify user interests and the effects of harvester Digital Public Library of America on Hunter Library digital collections (Biswas & Marchesoni, 2016, p. 30).

The Hunter Library relied upon certain tools to conduct this usage study. Google tool Query Explorer customized and directed Google Analytics to retrieve the aforementioned data. This data was collected within Excel, and then sorted via open source tool OpenRefine for ease of review (Biswas & Marchesoni, 2016, p. 23). This sorting enabled comparisons between categories of data, which shed certain insights about Hunter Library digital collections usage.

Tallied page views offered a picture of public interest (Biswas & Marchesoni, 2016, p. 23). Craft Revival collection statistics dominated those of the library's other collections with "[…] close to 42 percent of all item views and 53 percent of all item referrals" (Biswas & Marchesoni, 2016, p. 25). The Horace Kephart collection ranked second in usage with almost 12% of total views (Biswas & Marchesoni, 2016, p. 25). These percentages indicated the interest in the Craft Revival and Horace Kephart collections.

Search keywords supported interpretations of public interest in the Craft Revival and Horace Kephart collections. OpenRefine isolated each keyword and provided a count of instances in which a search was performed with each keyword. To account for variance amongst search terms which represent similar queries, OpenRefine has the capacity to sort similar keywords together regardless of the way that terms are spelled, or whether capitals are used (Biswas & Marchesoni, 2016, p. 25). This revealed search keywords to most commonly reference "[…] arts and crafts from the Western North Carolina region ("baskets," Indian masks," "Indian wood carving," "Cherokee pottery") […] Searches relating to the Horace Kephart collection ("horace kephart," "kephart knife") are also popular, explaining the fact that the Kephart collection […] scores highly in terms of item views (second) […]" (Biswas & Marchesoni, 2016, p. 27).

Meanwhile, referral data offered particularly significant insight for the Hunter Library's partnership with harvester DPLA. Search engines such as Google and Bing predominantly brought users to the digital collections. However, as a demonstration of DPLA impact on Hunter Library digital collections, referral data indicated that "Each of these collections gets 17 percent of referrals from the DPLA […] A trend seems also to show there is an increase in total referrals from DPLA per month the longer items are in DPLA […]" (Biswas & Marchesoni, 2016, p. 28).

A final insight was revealed via user geographic location information. This data indicated that many users accessed the Hunter Library digital collections while in North Carolina. This had to be reconciled with the irrelevance of the Craft Revival and Horace Kephart collections to Western Carolina University history (Biswas & Marchesoni, 2016, p. 30). It was determined that this data indicates usage of these digital collections by those with no affiliation with the university, which may be an effect of the niche role that these collections serve. Users returned to certain digital collections more than others, like Highlights from WCU, the collection with the largest percentage of instances of return visits (Biswas & Marchesoni, 2016, p. 30). Thus, this niche claim was supported as return visits indicate "[…] a loyalty to these collections" (Biswas & Marchesoni, 2016, p. 31).

These insights have implications for the Hunter Library moving forward. Each type of Google Analytics data corroborates the insights that other types reveal, offering assurances about the accuracy of each insight. Public interest in certain library collections supports the continued expansion of these collections. Indexing informed by common search keywords may promote findability and coherence of digital collections: "[…] social tagging […] is a powerful form of indexing because of "close connection with users and their language," as opposed to traditional indexing" (Biswas & Marchesoni, 2016, p. 31). Finally, user geographic information serves as

impetus for the library to reassess its designated community to consider the needs of its non-university users (Biswas & Marchesoni, 2016, p. 30).

Conclusion

Looking back at all that encompasses a museum digitized collections usage study, certain concepts stand out amongst the rest. A total grasp of usage requires capture of both reach and impact statistics. Various quantitative and qualitative methods exist to retrieve this usage data, which should be selected for implementation based upon the types of evidence that a museum hopes to capture, with data continuously gathered over an extended time frame to indicate patterns of usage. While these usage evaluation methods can be effective, certain variables may inhibit a museum's ability to implement these methods or to capture accurate data. Thus, reliance upon more than one quantitative and qualitative method and caution in interpreting resulting data at face value can paint a cleaner picture of usage. This glimpse of usage has profound implications for digitized collections sustainability as this data pinpoints user needs and interests. However, changes enacted in responding to usage data must be reconciled with any staff concerns about taking such actions. These themes collectively are illustrated in the two case studies described, which offer valuable insight for museums regarding usage study best practice.

The relevance of museum digitized collections usage studies cannot be overstated. The insight to be gained from such a study can inspire greater confidence in digitized collections resource allocation as museums are made aware of user actions and equipped to appropriately address user needs and interests. As websites compete for the attention of digital users, change which considers user needs and interests can ensure the relevance and usage of digitized collections. Thus, careful consideration of usage data will increase museum faith in digitized collections as an adequate response to public interest in digital content.

References

Ball, A. & Duke, M. (2015). How to track the impact of research data with metrics. *Digital Curation Centre*, 1-16. Retrieved from http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics

Bertacchini, E., & Morando, F. (2013). The future of museums in the digital age: New models for access to and use of digital collections. *International Journal of Arts Management, 15*(2), 60-72. Retrieved from http://www.jstor.org/stable/24587113

Biswas, P., & Marchesoni, J. (2016). Analyzing digital collections entrances: What gets used and why it matters. *Information Technology & Libraries, 35*(4), 19-34. Retrieved from https://ejournals.bc.edu/ojs/index.php/ital/article/view/9446/pdf

Breeding, M. (2009). Maximizing the impact of digital collections. *Computers in Libraries, 29*(4), 32-34.

Eccles, K. (n.d.). *What are Log Files?*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/41/what-are-log-files

Everett, C. (2017). GET FULL VALUE FROM DATA ANALYTICS PROGRAMMES. *Computer Weekly*, 16.

Farney, T. (2016). Getting the best Google Analytics data for your library. *Library Technology Reports, 52*(7), 5-8. Retrieved from https://journals.ala.org/index.php/ltr/article/view/6125/7902

Gerrard, D., Sykora, M., & Jackson, T. (2017). Social media analytics in museums: extracting expressions of inspiration. *Museum Management and Curatorship,* 32(3), 232-250. doi:10.1080/09647775.2017.1302815

Google (n.d.). *Google Scholar Citations*. Retrieved from https://scholar.google.com/intl/en/scholar/citations.html

Holter, E. (2017, July 25). *Measuring the impact of museum websites.* Retrieved from https://www.linkedin.com/pulse/measuring-impact-museum-websites-eric-holter

Hughes, L. M. (2012). Live and kicking: The impact and sustainability of digital collections in the humanities. In C. Mills, M. Pidd, and E. Ward (Eds.), *Proceedings of the Digital Humanities Congress 2012*. Retrieved from https://www.dhi.ac.uk/openbook/chapter/dhc-hughes

Hull, D. (2011). *Assessing the value and impact of museums*. Retrieved from http://www.niassembly.gov.uk/globalassets/Documents/RaISe/Publications/2011/Culture-Arts-Leisure/2911.pdf

Leon, S. (2012, September 17). The revenge of the yellow milkmaid: Cultural heritage institutions open up dataset of 20m+ items [Blog post]. Retrieved from https://blog.okfn.org/2012/09/17/the-revenge-of-the-yellow-milkmaid-cultural-heritage-institutions-open-up-dataset-of-20m-items/

Meyer, E. T. (n.d. a). *Bibliometrics: Enhancing the ability to track a project's scholarly impacts*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/53/bibliometrics-enhancing-ability-track-projects-scholarly-impacts

Meyer, E. T. (n.d. b). *Software tools for Analytics*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/44/software-tools-analytics

Meyer, E. T. (n.d. c). *Software tools for bibliometrics*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/49/software-tools-bibliometrics

Meyer, E. T. (n.d. d). *Software tools for content analysis*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/56/software-tools-content-analysis

Meyer, E. T. (n.d. e). *Surveys: General considerations*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/31/surveys-general-considerations

Meyer, E. T. (n.d. f). *Survey tools*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/28/survey-tools

Meyer, E. T. (n.d. g). *What are Analytics?*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/42/what-are-analytics

Meyer, E. T. (n.d. h). *What is Bibliometrics and Scientometrics?*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/48/what-bibliometrics-and-scientometrics

Meyer, E. T., & Eccles, K. (n.d.). *What are focus groups?*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/34/what-are-focus-groups

Mills, A. (2015). User impact on selection, digitization, and the development of digital special collections. *New Review of Academic Librarianship, 21*(2), 160-169. doi:10.1080/13614533.2015.1042117

Oxford Internet Institute, University of Oxford (n.d.). *About the toolkit*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/about-toolkit

Oxford Internet Institute, University of Oxford (2013). *Using Google Analytics*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/sites/microsites.oii.ox.ac.uk.tidsr/files/2013-06-15_GoogleAnalytics.pdf

Szajewski, M. (n.d.). Using Google Analytics data to expand discovery and use of digital archival content. *Practical Technology for Archives.* Retrieved from https://practicaltechnologyforarchives.org/issue1_szajewski/

Thelwall, M. (n.d.). *How do I set up log file analysis?*. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/kb/24/how-do-i-set-log-file-analysis

Verwayen, H., Arnoldus, M., & Kaufman, P. B. (2011). The problem of the yellow milkmaid: A business model perspective on open metadata. *Europeana*, 1-25. Retrieved from

https://pro.europeana.eu/files/Europeana_Professional/Publications/Whitepaper_2-The_Yellow_Milkmaid.pdf

Vogl, G., Zhou, Y., Draper, D., & Shelstad, M. (2016). Implementing Google Analytics for a consortial digital repository. *Library Technology Reports, 52*(7), 31-37. Retrieved from https://journals.ala.org/index.php/ltr/article/view/6130/7912